aa-tRNA competition is crucial for the effective translation efficiency

Wenjun Xia[a], Jinzhi Lei[b]

[a]*Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University, Beijing 100084, China*
[b]*MOE Key Laboratory of Bioinformatics, Zhou Pei-Yuan Center for Applied Mathematics, Tsinghua University, Beijing 100084, China*

Abstract

Translation is a central biological process by which proteins are synthesized from genetic information contained within mRNAs. Here we study the kinetics of translation at molecular level through a stochastic simulation model. The model explicitly include RNA sequences, ribosome dynamics, tRNA pool and biochemical reactions in the translation elongation. The results show that the translation efficiency is mainly limited by the available ribosome number, translation initiation and the translation elongation time. The elongation time is log-normal distribution with mean and variance determined by both the codon saturation and the process of aa-tRNA selection at each codon binding. Moreover, our simulations show that the translation accuracy exponentially decreases with the sequence length. These results suggest that aa-tRNA competition is crucial for both translation elongation, translation efficiency and the accuracy, which in turn determined the effective protein production rate of correct proteins. Our results improve the dynamical equation of protein production with a delay differential equation which is dependent on sequence informations through both the effective production rate and the distribution of elongation time.

Keywords: RNA translation, efficiency, gene regulation network, ncRNA

1. Introduction

Translation is a central biological process by which genetic information contained within mRNAs is interpreted to generate proteins. Ribosomes provide the environment for all activities of the translation process, such as the formation of the initiation complex, the elongation of the translation process by which the ribosome moves along the mRNA sequence, and the dissociation of the ribosome from the mRNA. Protein synthesis is principally regulated at the initiation stage and hence the protein production rate is mainly limited by the availability of free ribosomes [12, 20]. During translations, the ribosome selects matching aminoacylated tRNA (aa-tRNA) to the mRNA codon from a bulk of non-matching tRNAs, and the reaction rate constants can show 350-fold difference in the stability of cognate and near-cognate codon-anticodon complexes [8]. Hence, in addition to the initiation stage, the translation efficiency is also affected by the mRNA sequence and the competition between cognate and near-cognate tRNAs [6, 16, 24]. Global understanding of how ribosome number, mRNA sequence, and tRNA pool combine to control the translation kinetics has been an interesting topic in recent years for its potential impacts on the biogenesis and synthesis biology [9, 13, 15, 20].

Computational models have been developed to investigate details of the translation kinetics and to explore the main factors that affect the translation efficiency, such as codon bias, tRNA and ribosome competition, ribosome queuing, codon order [2, 3, 6, 14, 19, 20, 21]. In these models, status of all ribosomes and tRNAs along a mRNA are tracked in continuous time. Translation initiation and the availability of free ribosomes were highlighted in previous studies [3, 20, 21]. In [21], it was found that the varying of translation efficiency were caused by very short times of translation initiation. Through a model that tracks all ribosomes, tRNAs and mRNAs in a cell, it was concluded that the protein production in healthy yeast cells was typically limited by the availability of free ribosomes, however the protein production under stress was rescued by reducing the initiation or elongation rates [20] . Codon bias of a mRNA sequence is an important factor that may affect the the translation efficiency due to competitions of tRNAs [2, 3, 6]. A study of *S. cerevisiae* genome suggest that tRNA diffusion away from the ribosome is slower than translation, and hence codon correlation in a sequence can accelerate translation because the same tRNA can be used by nearby codons [2]. In the elongation process, a cognate, near-cognate, or non-cognate tRNA may attempt to bind to the A site of a ribosome. A study based on a computation model with detailed tRNA pool composition shows that the competition between near-cognate and cognate tRNAs is a key factor that determines the translation rate [6]. Another study by a mean-field model of translation in *S. cerevisiae* shows that the competition for ribosomes, rather than tRNAs, limits global translation [3]. Ribosome collisions can also reduce the translation efficiency according to a model of stochastic translation process for *E. coli lacZ* mRNA as a traffic problem [14]. From the point of view of evolution, the mechanism for controlling the efficiency of protein translation was evolutionarily conserved according to a calculation on adaptation between coding sequences and the tRNA pool [23]. Moreover, using a nested model of protein translation and population genetics to the genomic of *S. cerevisiae*, it was suggested that the codon usage bias of genes can be explained by the evolution through the selections for efficient ribosomal usage, genetic drift, and biased mutation, and the selection for efficient

ribosome usage is a central force in shaping codon usage at the genomic scale [19].

Despite extensive studies, much details of how translation is controlled by mRNA sequences and cellular environment remain exclusive. Both the number of available free ribosomes and the codon orders are important for translation efficiency, however how various factors combine to determine the translation efficiency is not clearly formulated. Since a codon is bound by a near-cognate tRNA, proteins with mismatched amino acids can be produced in translations. Hence, the translation accuracy may be dependent on the codon usage of a sequence and the composition of tRNAs, but little result about the dependence is known to the best of our knowledge. The relation of how the timing of ribosome elongation stage depends on a sequence and the tRNA pool is closely related to the modeling of genetic network dynamics in which the elongation time associates with the time-delay in dynamical equations [22, 26, 27], but how elongation time is formulated remains mystery.

In this paper the translation kinetics is considered through a stochastic computation model with detailed reactions of the ribosome dynamics. In our study, several factors including the coding sequence, ribosomes, and the composition of tRNA tool were modeled to investigate how the translation efficiency, accuracy, and elongation time are determined. Moreover, translation dynamics of various mRNA sequences (yeast and human, coding and non-coding mRNAs) were studied to try to clarify whether or not the sequence is important for the translation efficiency and the translation accuracy. Our results show that the translation efficiency is mainly limited by the number of the available ribosomes, translation initiation and the elongation time of translation, and the elongation time is log-normal distribution with mean and variance of the logarithm of the elongation time dependent on the sequence through aa-tRNA usages. Moreover, the translation accuracy exponentially decrease with the sequence length. These results provide more detailed understanding of the translation processes, and can improve the mathematical modeling of protein production in gene regulation network dynamics.

2. Model and methods

2.1. Model description

Fig. 1 illustrates our model of ribosome kinetics in translation presented in [6][1]. We summarize the model description below and refer [6] for details.

Translation of a protein begins from the initiation stage by which the start codon (AUG site) of the mRNA sequence is occupied by a ribosome, and the peptide between the first two amino acids are formed, with corresponding aa-tRNAs binding to the E and P sites of the ribosome, respectively. During the elongation, each move of the ribosome includes 9 steps as shown by Fig. 1: initial binding of aa-tRNA, codon recognition, GTPase activation, GTP hydrolysis, EF-Tu conformation change, rejection, accommodation, peptidyl transfer, and
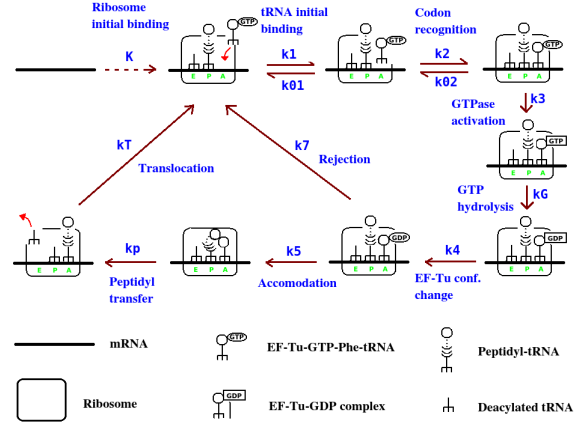
[1]See http://v.youku.com/v_show/id_XNzMxNzEwNjg0.html for an animation of translation. Kindly provided by Prof. Ada Yonath.

Figure 1: Kinetic scheme of RNA translation. Re-drew from [6].

Table 1: Values of kinetic rate constants (s^{-1}) (refer to [6])

Parameters	Values	Cognate	Near-cognate	Non-cognate
K	0.03	-	-	-
k1	-	140	140	2000
k01	-	85	85	-
k2	-	190	190	-
k02	-	0.23	80	-
k3	-	260	0.4	-
kG	-	1000	1000	-
k4	-	1000	1000	-
k5	-	1000	60	-
k7	-	60	1000	-
kp	-	200	200	-
kT	-	20	20	-

translocation. For each codon on the mRNA sequence, tRNAs in the tRNA pool are divided into three types: cognate, near-cognate, and non-cognate, as listed in [6]. All aa-tRNAs can attempt to bind to the A site of the ribosome according to the match between codon and anticodon [8], however only cognate and near-cognate aa-tRNAs can go through the step of peptide formation, non-cognate aa-tRNAs are rejected by codon recognition. Cognate aa-tRNAs give the correct amino acid following the genetic code, while near-cognate tRNAs often bring incorrect amino acids and yield a defect protein. Reactions rates are different for cognate and near-cognate tRNAs, which have been reported at [8, 18] and are given by Table 1 in our simulations. We note that near-cognate aa-tRNAs are more likely to be rejected at both steps of codon recognition and rejection. Therewith, the competition between cognate and near-cognate tRNAs may be crucial for both the fidelity of peptide synthesis and translation efficiency [6, 8]. After peptidyl transfer, the E site aa-tRNA is released and the ribosome move forward a codon with the A site free waiting for the next move. Translation of a polypeptide stops when the ribosome reaches a stop codon (UAG/UAA/UGA), where the polypeptide is released and the ribosome drops off from the mRNA. One ribosome can synthesize only one polypeptide at a time, and each mRNA can be translated simultaneously by multiple ribosomes. The multiple ribosomes forms a queue along the mRNA, with a safe distance of at least 10 codons between two ribosomes [14, 20].

Table 2: tRNA pool composition (refer to [5, 6]). Also refer [6] for the anti-codons for the tRNAs.

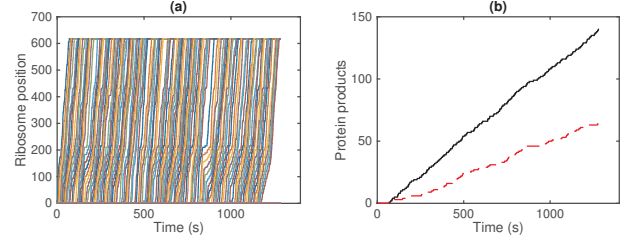| tRNA	Molecules/cell	tRNA	Molecules/cell	tRNA	Molecules/cell
Ala1	3250	His	639	Pro3	581
Ala2	617	Ile1	1737	Sec	219
Arg2	4752	Ile2	1737	Ser1	1296
Arg3	639	Leu1	4470	Ser2	344
Arg4	867	Leu2	943	Ser3	1408
Arg5	420	Leu3	666	Ser5	764
Asn	1193	Leu4	1913	Thr1	104
Asp1	2396	Leu5	1031	Thr2	541
Cys	1587	Lys	1924	Thr3	1095
Gln1	764	Met f1	1211	Thr4	916
Gln2	881	Met f2	715	Trp	943
Glu2	4717	Met m	706	Tyr1	769
Gly1	1068	Phe	1037	Tyr2	1261
Gly2	1068	Pro1	900	Val1	3840
Gly3	4359	Pro2	720	Val2A	630
				Val2B	635

Figure 2: Translation kinetics of a single mRNA sequence. (a) Positions of each ribosome on the sequence. (b) Numbers of protein products. Black solid line for all protein products, red dashed line to correctly translated proteins (no incorrect amino acid added by near-cognate aa-tRNAs). Here the sample sequence is the gene YAL003W from the SGD yeast coding sequences, with sequence length $L = 621nt$. Parameters are $R = 20$, $F = 0.03$ and other parameters refer Table 1.

2.2. Numerical scheme

The translation process with multiple ribosomes was modeled with the stochastic simulation algorithm (SSA) [7], which includes the following reactions:

1. binding of a ribosome to the start codon if the first 10 codons are not occupied by ribosomes;
2. binding of an aa-tRNA from the tRNA pool to the A site of an unoccupied ribosome;
3. reactions of codon recognition, energy transformation, and peptide formation;
4. releasing of the tRNA from the E site of a ribosome;
5. translocation of the ribosome to the next codons if the safety condition is satisfied;
6. dropping off of the ribosome once the stop codon is reached.

Kinetics parameters are given by Table 1, which refer to [6]. The tRNA pool compositions referred the total number of each tRNA in a yeast cell from [5, 6] and are given by Table 2. In simulations, to mimic the effects of available tRNAs for each single mRNA translation, we used a factor F ($0 < F \leq 1$) to all tRNA numbers to adjust the changes in the numbers of available tRNAs. For the anti-codon of each tRNA and the cognate, near-cognate, and non-cognate for each codon, refer [6] for details.

It has been shown that the availability of free ribosomes is an important limitation for the translation efficiency [20]. Here we introduced a parameter R for the maximum number of available ribosomes can be used for a single sequence translation. We note that a ribosome can be re-used after it was released from the stop codon.

An example of translation kinetics obtained from our simulation is given at Fig. 2, which shows that the ribosomes sequencing along the mRNA, and the number of protein production increases linearly with the translation time. The average translation rate (amino acids per second) in our simulation is of a order of 10, in good agreement with the experimental observations [17]. These suggest well defined translation efficiency, elongation time, and accuracy of a translation as given below.

2.3. Translation efficiency, elongation time, and accuracy

To quantify the translation process, we consider the translation efficiency for the protein production rate, the elongation time for move kinetics of each individual ribosome, and the translation accuracy for the fidelity of translation. The *translation efficiency* (TE) is defined as the average slope of the increasing of protein production number with the translation time. The *elongation time* of each ribosome is given by the time period from the binding of a ribosome to the start codon to its dropping off from a stop codon. The *elongation time per codon* (ETC), the average time for a ribosome to move one codon, is often used to describe the translation kinetics. The elongation time is given by $ETC \times L/3$, where L is the length (in *nt*) of a mRNA sequence. Since a protein product may contain mismatched amino acids due to the binding of near-cognate aa-tRNAs with the mRNA, it was possible to have incorrect protein products in the translation. Hence, the ratio of correct proteins in all protein produces gives the *translation accuracy*.

3. Results

3.1. Translation elongation time is log-normal distribution and sequence dependent

The elongation time measures how long it takes a ribosome to finish the translation of a protein, which corresponds to the delay of translation in modeling the dynamics of gene regulation networks through delay differential equations [26, 27]. The production of proteins can be described by translation efficiency α and mRNA number $M(t)$ through a delay differential equation of form

$$\frac{dP}{dt} = \alpha \int_0^{+\infty} M(t - \tau)\rho(\tau)d\tau, \tag{1}$$

where τ represents the elongation time, with distribution density $\rho(\tau)$.

To obtain the formulation of the distribution density $\rho(\tau)$, we calculated the elongation time per codon (ETC) in the translation of YAL003W (here we note $\tau = ETC \times L/3$). The distribution density is showed at Fig 3. The density function was well fitted by log-normal distribution

$$\ln \mathcal{N}(\mu, \sigma^2) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, \quad x > 0. \tag{2}$$

Here the shape parameters μ and σ are taken so that the logarithm of ETC has mean μ and variance σ^2. Let $n = L/3$ be

the number of amino acids in a protein product, the density function $\rho(\tau)$ of elongation time is given from the log-normal distribution Eq. 2 as

$$\rho(\tau) = \frac{1}{\tau\sigma\sqrt{2\pi}}e^{-\frac{(\ln\tau-\ln n-\mu)^2}{2\sigma^2}}, \qquad (3)$$

and the average elongation time is

$$\bar{\tau} = \int_0^{+\infty}\tau\rho(\tau)d\tau = ne^{\mu+\sigma^2/2}. \qquad (4)$$

In the next section we show that the translation efficiency is dependent on the average elongation time, which refines our dynamical equation for protein production.
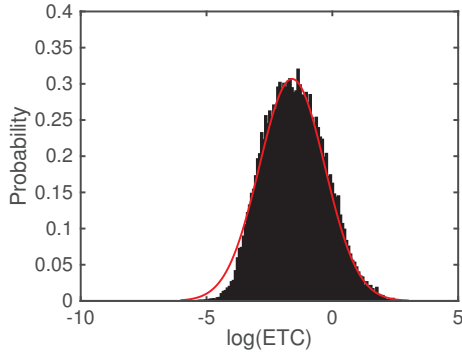
Figure 3: Distribution of the elongation time per codon in the translation of YAL003W. All parameters are the same at Fig. 2. Red curve is the fit with normal distribution $\ln \mathcal{N}(-1.6, 1.69)$.

Each move of a ribosome consists of several chemical reactions shown by Fig. 1, including selections of cognate or near-cognate aa-tRNA from the tRNA pool, and a step forward if the safety condition is satisfied. To investigate how the ETC depends on the mRNA sequence and translation kinetics, we examined translations of a set of 1000 sequences from yeast coding genes with length L varies from 51 to 1995 nt (17 to 665 codons). To measure the tRNA usage of each sequence, we calculated the average fraction of cognate, near-cognate, and non-cognate tRNA along the sequence, which are defined as

$$F_\nu = \frac{1}{L/3}\sum_{i=1}^{L/3}\frac{n_{i,\nu}}{\text{Total tRNA number}}, \ \nu = \text{cog, near, non,} \quad (5)$$

here F_ν ($\nu = $ cog, near, non) measures the average tRNA usage of cognate, near-cognate, and non-cognate tRNAs, respectively. The summation is taken over all codons, and $n_{i,\nu}$ is the number of tRNAs of type ν for codon i along the mRNA sequence.

Fig. 4 shows the dependence of the mean (μ) and variance (σ^2) of the logarithm of ETC with tRNA usages. Results suggest that the mean decreases with the cognate tRNA usage F_{cog}, increases with the near-cognate tRNA usage F_{near}, and has no correlation with the non-cognate tRNA usage F_{non}, while the variance is not dependent on either F_{cog} or F_{near}, but weakly decreases with F_{non}. These results suggest that the competition of near-cognate tRNAs tends to increase the elongation time,

while the competition of non-cognate tRNAs has only little affects to the elongation time. Moreover, Fig. 4 suggests typical parameters for the distribution of the ETC of yeast coding gene translation are $\mu \approx -1.5$ and $\sigma^2 \approx 1.4$ (refer to Eq. 2). Our simulations suggest no obvious dependence of ETC with sequence length L (data not shown).
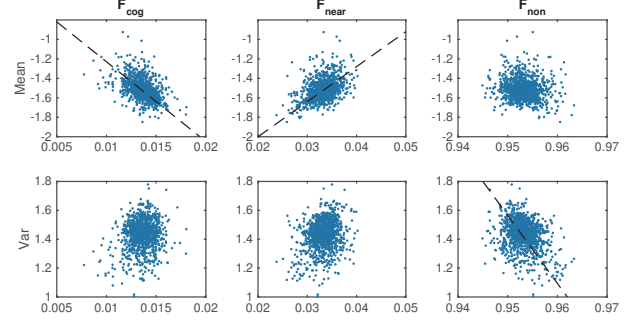
Figure 4: Dependence of the ETC of yeast coding sequences on tRNA usages. Dots show the mean (upper panel) and variance (bottom panel) of the logarithm of ETC with cognate tRNA usage F_{cog}, near-cognate tRNA usage F_{near} and non-cognate tRNA usage F_{non}, respectively. Dashed lines show the linear fitting. Simulations of 1000 yeast coding sequences are shown, each dot corresponds to one sequence. All parameters are the same as Fig. 2.

The investigate how the available ribosomes number R affects the elongation time, we changed the value R to calculate the dependence of ETC. Results showed that both mean and variance of the logarithm of ETC are dependent on R nonlinearly: mostly independent to R when R is either small or large, and an obvious increasing dependence when R takes intermediate values (Fig. 5a). A possible reason for the increasing of ETC is the traffic jam due to codon occupation. Fig. 5b shows that the average ribosome distance obviously decreases with R at the intermediate region, and approaches a minimum distance (the safe distance of 10 codons) when R is large. These results reveal that the increasing dependence of the elongation time with ribosome number R ($10 < R < 30$) is due to the increasing of traffics jam in translation kinetics.
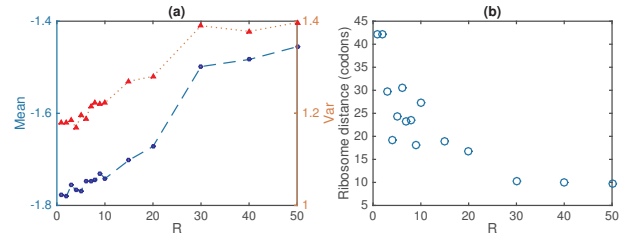
Figure 5: Dependence of the elongation time with the available ribosomes number R. (a) Average ETC versus R. (b) Ribosome distance (in codons) versus R. Sequence and parameters are the same as Fig. 2.

In the above calculations, the total number of tRNAs was fixed. To further examine how the number of total tRNAs affects the elongation time, we varied the factor F from 0.03 to 1 to calculate the dependence of ETC. Results showed that both mean (μ) and variance (σ^2) of the logarithm of ETC are decreasing with F for small F, and nearly unchanged when $F > 0.5$ (Fig. 6). Biologically these dependences are obvious

because it takes longer time to select a cognate or near-cognate when there are no enough tRNAs.
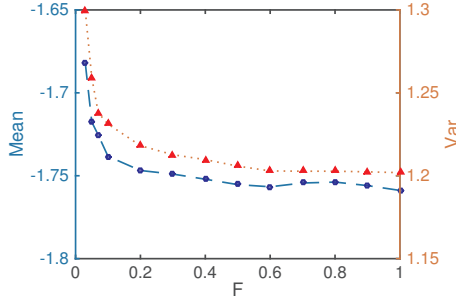

Figure 6: Dependence of the ETC with total tRNAs number represented by the factor F. The mean (left hand ordinate, blue circles connected with a dashed line) and variance (right hand ordinate, red triangles connected with a dotted line) of the logarithm of ETC are shown as a function of the factor F. Sequence and parameters are the same as Fig. 2.

3.2. Translation efficiency is mainly dependent on the elongation time and available ribosomes number

In [20, 21], it has been shown that the varying of translation efficiency were caused by translation initiation, and the availability of free ribosomes was a typical rate limit of translation. To investigate how the translation efficiency depends on the translation kinetics and mRNA sequences, we constructed a model to track the dynamics of available ribosomes.

Consider a mRNA with n codons ($n = L/3$). Let R be the number of available ribosomes, $x_i(t)$ ($i = 1, \cdots, n$) the number of ribosomes at the ith codon at time t, and $x_0(t)$ the number of free ribosomes. Kinetics of a ribosome in translation is a combination of initiation in a rate K, elongation per codon in a rate c and termination in a rate K_T. Therefore, the dynamics of x_i can be expressed by the following differential equations model

$$
\begin{cases}
\dfrac{dx_0}{dt} &= K_T x_n - K x_0 \\
\dfrac{dx_1}{dt} &= K x_0 - c x_1 \\
\dfrac{dx_i}{dt} &= c(x_{i-1} - x_i) \qquad i = 2, 3, \cdots, n-1 \\
\dfrac{dx_n}{dt} &= c x_{n-1} - K_T x_n.
\end{cases}
\tag{6}
$$

The protein production rate is proportional to x_n. Here we note

$$
0 \le x_0 \le R, \quad 0 \le x_i \le 1 \quad (i = 1, 2, \cdots, n), \tag{7}
$$

and

$$
\sum_{i=0}^{n} x_i = R. \tag{8}
$$

When $K_T > c$ and R is small, Eq. 6 has a stable equilibrium state which gives

$$
x_n = \frac{R}{K_T \dfrac{n-1}{c} + 1 + \dfrac{K_T}{K}}. \tag{9}
$$

Hence, let $\bar{\tau} = \dfrac{n-1}{c}$ be approximate to the elongation time (here we note $1/c$ corresponds to the average of ETC), the translation efficiency satisfies

$$
TE \propto \frac{RK}{K\bar{\tau} + 1 + K/K_T}. \tag{10}
$$

When R is large so that all codons are occupied, the translation efficiency is mainly determined by the elongation time so that $TE \propto 1/\bar{\tau}$. Hence, taking account Eq. 10, the translation efficiency can be approximated as

$$
TE \propto \frac{K \min\{R, R_{\max}\}}{K\bar{\tau} + 1 + K/K_T}, \tag{11}
$$

where $R_{\max}$ is the number of available ribosomes to saturate all codons. We take $R_{\max} = n/10$ in our simulations, which is consistent with Fig. 5. We note that $\bar{\tau}$ is dependent on R according to the above discussions, hence the relation Eq. 11 suggests the following dependence of translation efficiency on the ribosome number R: linearly increases when R is small, independent to R when R is large, and nonlinear dependence through the elongation time $\bar{\tau}$ when R takes intermediate values. These results are in agree with our numerical simulations (Fig. 7).
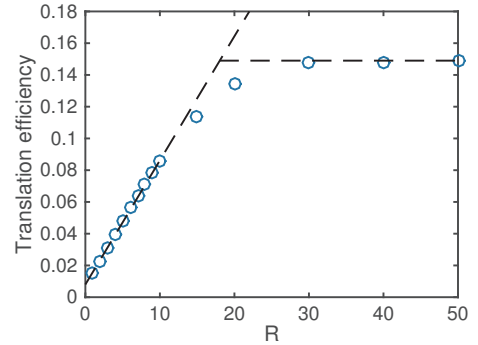

Figure 7: Dependence of translation efficiency with the maximum number of available ribosomes R. Dashed lines show the two-phase dependence following Eq. 11. Sequence and parameters are the same as Fig. 2.

The result Eq. 11 supports the previous findings that translation initiation and ribosome number are rate limits of protein production. Moreover, the translation efficiency decreases with the elongation time, which shows the dependence of protein production with the mRNA sequence through the elongation dynamics.

Since the average elongation time $\bar{\tau}$ is proportional to the protein length n, Eq. 11 suggests that the translation efficiency depends on the protein length n through a Michaelis-Menten function. Fig. 8a shows translation efficiency versus sequence length for yeast coding sequences with different length. The translation efficiency is well fitted by a Michaelis-Menten function, in agreement with our theoretical conclusion Eq. 11.

To further investigate the sensitivity of translation efficiency with the changes in parameters, we increased or decreased each of the parameters in Table 1 and examined the changes in the
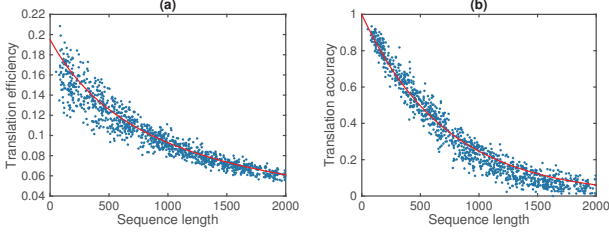
5

Figure 8: Translation kinetics. (a) Translation efficiency versus sequence length for 1000 yeast coding sequences. Red line shows the fitting with $TE = \dfrac{0.195}{1 + 0.0033n}$. (b) Translation accuracy versus sequence length for 1000 yeast coding genes. Red line shows the fitting with $e^{-0.0042n}$. Here $n = L/3$ is the protein chain length. Data obtained from the simulation at Fig. 4.

translation efficiency. The results showed that the translation efficiency is sensitive with the changes in k02 (or ke02 for near-cognate tRNA), k01, k1, k2, which correspond to the process of aa-tRNA selection. The translation initiation K is also important for the translation efficiency, as we have seen from Eq. 11. Changes in other parameters led to minor changes in the translation efficiency. These results indicate that the steps of aa-tRNA selection are crucial for the translation efficiency through their effects to the elongation time, and changes in the steps of peptide formation have minor effects to the translation efficiency.
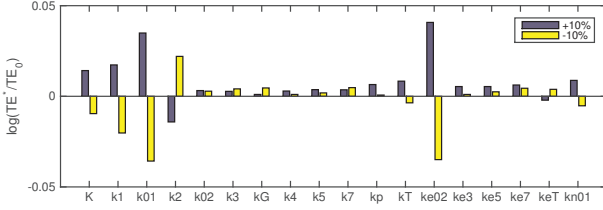

Figure 9: Sensitivity analysis of the translation efficiency. Bars show changes in the logarithm of translation efficiencies induced by changes in a single parameter $\ln(TE^*/TE_0)$, where TE^* and TE_0 represents the TE for modified and default parameters, respectively. Blue bars correspond to increasing of a parameter by 10%, and yellow bars correspond to decreasing of a parameter by 10%. For parameters refer to Table 1, the parameters ke02, ke3, ke5, ke7, keT for values of k02, k3, k5, k7, kT of near-cognate tRNAs (second column in Table 1), respectively, and kn01 for the parameter k01 for non-cognate tRNAs (third column in Table 1. Sequence and defaults parameters are same as Fig. 2.

3.3. Translation accuracy decreases exponentially with sequence lengths

In translations, protein products may contain mismatched amino acids when a near-cognate aa-tRNA is selected and successfully form a peptide. Hence the translation accuracy, fraction of correct protein products, should be exponentially decay with the chain length, the decay rate is associated with the probability of selecting a near-cognate aa-tRNA at each step. Fig. 8b shows the translation accuracy versus sequence length, which is well fitted with an exponential function.

In living cells, abnormal proteins are usually degraded quickly so that the intracellular amino acids can be recycled efficiently. Hence, only correctly translated proteins are relevant in modeling the dynamics of gene regulation networks. This

yields a factor by translation accuracy in the production rate of normal proteins in the equation Eq. 1.

The above discussions suggest a more refined equation for effective protein production Eq. 1 with $\rho(\tau)$ given by log-normal distribution Eq. 3, and the effective translation efficiency α is given by

$$\alpha = \frac{ae^{-cn}}{1 + bne^{\mu+\sigma^2/2}}, \qquad (12)$$

where parameters a, b depend on available ribosomes number, translation initiation and termination, and c relates with the composition of the tRNA pool. A crucial refinement of Eq. 12 is the dependence of protein chain length n, and other parameters are somehow universal for differential proteins, except weak dependence of μ and σ^2 with sequences as shown by Fig. 4, under certain cellular conditions.

3.4. Translation kinetics with sequence dependences

A motivation of this study was trying to examine whether there are distinct dynamics for coding and non-coding RNA sequences in the translation. We have shown that the translation efficiency depends on the mRNA sequence through the elongation time, and the mean and variance of the elongation time per codon are dependent on the sequence through the aa-tRNA usages. A study of ribosome occupancy showed that many large noncoding RNAs are bound by ribosomes and hence are possible to be translated into proteins [10, 11]. To investigate the translation kinetics of coding and noncoding RNAs, we applied the model simulation to yeast coding RNA, yeast noncoding RNA, human coding RNA and human noncoding RNA. In each sample, 500 sequences with lengths between $200nt$ and $1000nt$ were selected, however most of the noncoding RNA has reading frames with lengths less than $300nt$, in agreement with the observations in [4]. Simulations showed that the previous results are qualitatively held for different samples. Fig. 10 shows the distributions of mean (μ) and variance (σ^2) of the logarithm of the ETC for each set of the simulations (The average of μ and σ^2 for each ample are given by the table), which are crucial parameters in the density function Eq. 2. From Fig. 10, we have the following observations: coding RNAs (both yeast and human) have similar distribution in μ and σ^2; noncoding RNAs have smaller variance σ^2 comparing with the corresponding coding RNAs. These results reveal distinct translation kinetics statistically between coding and noncoding RNAs, waiting biological significance of the findings to be discovered.

4. Discussions

We have applied stochastic simulation to study the translation kinetics at molecular level. In the model, RNA sequences, ribosome dynamics, tRNA pool and biochemical reactions in the elongation steps were included. Simulations showed that during translations the ETC satisfied log-normal distribution (Fig. 3), and is mainly determined by both codon saturation (Fig. 5) and the steps of aa-tRNA selection (Fig. 9). In the tRNA selection, the relative numbers of near-cognate to cognate aa-tRNAs are crucial for the elongation of a ribosome, and
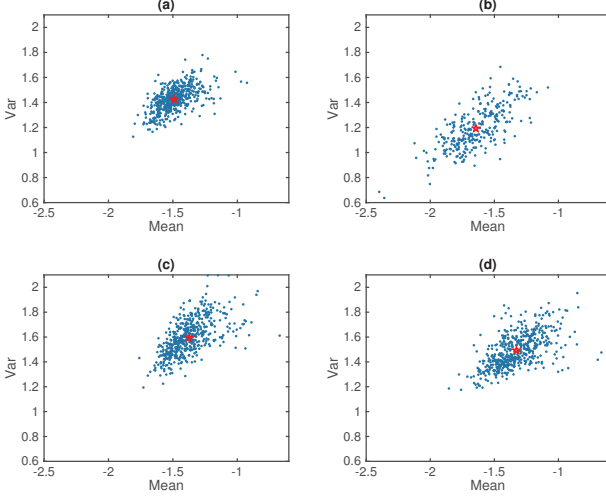
	Mean (μ)	Variance (σ^2)
Yeast coding	-1.4895	1.4256
Yeast non-coding	-1.6410	1.1948
Human coding	-1.3773	1.5981
Human non-coding	-1.3295	1.4885

Figure 10: *ETC* of the translations for different samples. Distributions of Mean and variance of the logarithm of *ETC* for yeast coding RNAs (a), yeast noncoding RNAs (b), human coding RNAs (c) and human noncoding RNAs (d). Here the results of 500 random sequences with length $200nt < L < 1000nt$ for each sample are shown. Red stars show the average values for each sample, values are given by the table. Parameters are $R = 20, F = 0.03$, and other parameters are referred to Table 1.

hence the mean value of the logarithm of *ETC* is dependent on the tRNA usages defined by F_{cog} and F_{near} in this paper (Fig. 4). In the log-normal distribution Eq. 2, the mean μ and variance σ^2 are important for the density function of the elongation time. We showed that these two parameters are slightly different for coding and noncoding RNAs for both yeast and human samples (Fig. 10). On average, noncoding RNAs have smaller variance than coding RNAs. A simple model of ribosome dynamics revealed that the translation efficiency is mainly determined by the number of available ribosomes, translation initiation and the elongation time, and the translation efficiency depends on the elongation time through a Michaelis-Menten function. The translation efficiency increases with the available ribosomes number when the number is small, however is insensitive with the ribosome number when the number is large enough to saturate all codons. These results are further confirmed by our simulations. Moreover, the translation accuracy decreases exponentially with sequence lengths. These results suggest an improvement for the effective protein production when we are modeling gene expressions in gene regulation networks.

In modeling gene expressions the protein production is described by a delay differential equation of form Eq. 1 that depends on the translation efficiency α and the distribution $\rho(\tau)$ of elongation time. This study showed that the effective production of correct proteins can be expressed as

$$\alpha = \frac{ae^{-cn}}{1 + bne^{\mu+\sigma^2/2}}, \tag{13}$$

here n is the protein chain length (number of amino acids), μ and σ^2 are mean and variance of the logarithm of the *ETC*, respectively, parameters a and b are dependent on available ribosomes number, translation initiation and termination, and c relates with the composition of the tRNA pool. The distribution of the elongation time is formulated as

$$\rho(\tau) = \frac{1}{\tau\sigma\sqrt{2\pi}}e^{-\frac{(\ln\tau-\ln n-\mu)^2}{2\sigma^2}}. \tag{14}$$

Hence, the protein production equation Eq. 1 can be rewritten as

$$\frac{dP}{dt} = \frac{ae^{-cn}}{1 + bne^{\mu+\sigma^2/2}}\int_0^{+\infty}M(t-\tau)\frac{1}{\tau\sigma\sqrt{2\pi}}e^{-\frac{(\ln\tau-\ln n-\mu)^2}{2\sigma^2}}d\tau, \tag{15}$$

where $M(t)$ is the number of mRNAs at time t. In this equation, the protein chain length n is explicitly included. Moreover, the sequence information are implicitly included in the parameter τ and σ^2 which are mainly determined by the process of aatRNA selection in each step of ribosome movement. Yet other parameters are somehow universal under given cellular conditions. A direct conclusion from Eq. 15 is the extreme low effective production rates of long proteins, which is because of the long elongation time and low translation accuracy for these long chain molecules. This is in consistent with biological observations that many transcription factors are small proteins with high production rates (many of them have high degradation rates as well) [1], and many structural proteins (*e.g.* fibers) and transport proteins (*e.g.* membrane proteins) are large proteins with low production rates (these proteins are mostly very stable) [25]. Hence, this study provides insightful details for the known observations, and is valuable in further works of whole cell modeling.

Data Resources

RNA sequences are downloaded from available databases:

- Yeast coding RNAs from SGD (http://downloads.yeastgenome.org/sequence/S288C_reference/ orf_dna/orf_coding.fasta.gz).

- Yeast noncoding RNAs from SGD (http://downloads.yeastgenome.org/sequence/S288C_reference/ rna/rna_coding.fasta.gz).

- Human coding RNAs from Ensembl Genome Browser (http://useast.ensembl.org/biomart/martview/8a921ac1ac4642b 07708af32f2339655).

- Human noncoding RNAs from Genecode19 (ftp://ftp.sanger.ac.uk/pub/gencode/Gencode_human/release_19 /gencode.v19.lncRNA_transcripts.fa.gz).

References

[1] Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M., & Teichmann, S. A. 2004. Structure and evolution of transcriptional regulatory networks. *Curr. Opin. Struct. Biol.*, **14**, 283–291.

[2] Cannarozzi, G., Schraudolph, N. N., Faty, M., von Rohr, P., Friberg, M. T., Roth, A. C., Gonnet, P., Gonnet, G., & Barral, Y. 2010. A role for codon order in translation dynamics. *Cell*, **141**(2), 355–367.

[3] Chu, D., Barnes, D. J., & von der Haar, T. 2011. The role of tRNA and ribosome competition in coupling the expression of different mRNAs in Saccharomyces cerevisiae. *Nucleic. Acids. Res.*, **39**, 6705–6714.

[4] Core, L. J., Martins, A. L., Danko, C. G., Waters, C. T., Siepel, A., & Lis, J. T. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat. Genet.*, **46**(12), 1311–1320.

[5] Dong, H., Nilsson, L., & Kurland, C. G. 1996. Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates. *J. Mol. Biol.*, **260**, 649–663.

[6] Fluitt, A., Pienaar, E., & Viljoen, H. 2007. Ribosome kinetics and aa-tRNA competition determine rate and fidelity of peptide synthesis. *Comput. Biol. Chem.*, **31**(5-6), 335–346.

[7] Gilliespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**(25), 2340–2361.

[8] Gromadski, K. B., & Rodnina, M. V. 2004. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol. Cell*, **13**(2), 191–200.

[9] Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S., & Lander, E. S. 2013. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell*, **154**(1), 240–251.

[10] Ingolia, N. T., Ghaemmaghami, S., Newman, J. R., & Weissman, J. S. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, **324**(5924), 218–223.

[11] Ingolia, N. T., Lareau, L. F., & Weissman, J. S. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*, **147**(4), 789–802.

[12] Jackson, R. J., Hellen, C. U., & Pestova, T. V. 2010. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.*, **11**(2), 113–127.

[13] Mao, Y., Liu, H., Liu, Y., & Tao, S. 2014. Deciphering the rules by which dynamics of mRNA secondary structure affect translation efficiency in Saccharomyces cerevisiae. *Nucleic. Acids. Res.*, **42**(8), 4813–4822.

[14] Mitarai, N., Sneppen, K., & Pedersen, S. 2008. Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization. *J. Mol. Biol.*, **382**(1), 236–245.

[15] Ninio, J. 2012. Ribosomal Kinetics and Accuracy: Sequence Engineering to the Rescue. *J. Mol. Biol.*, **422**(3), 325–327.

[16] Plotkin, J. B., & Kudla, G. 2010. Synonymous but not the same: the causes and consequences of codon bias. *Nat. Rev. Genet.*, **12**(1), 32–42.

[17] Proshkin, S., Rahmouni, A. R., Mironov, A., & Nudler, E. 1996. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *J. Mol. Biol.*, **260**, 649–663.

[18] Savelsbergh, A., Katunin, V., Mohr, D., Peske, F., Rodnina, M., & Wintermeyer, W. 2003. An elongation factor G-induced ribosome rearrangement precedes tRNA-mRNA translocation. *Mol. Cell*, **11**, 1517–1523.

[19] Shah, P., & Gilchrist, M. A. 2011. Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc. Natl. Acad. Sci. USA*, **108**(25), 10231–10236.

[20] Shah, P., Ding, Y., Niemczyk, M., Kudla, G., & Plotkin, J. B. 2013. Rate-limiting steps in yeast protein translation. *Cell*, **153**(7), 1589–1601.

[21] Siwiak, M., & Zielenkiewicz, P. 2010. A comprehensive, quantitative, and genome-wide model of translation. *PLoS Comput. Biol.*, **6**(7), e1000865.

[22] Tian, T., Burrage, K., Burrage, P. M., & Carletti, M. 2007. Stochastic delay differential equations for genetic regulatory networks. *J. Comput. Appl. Math.*, **205**(2), 696–707.

[23] Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., & Pilpel, Y. 2010a. An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, **141**(2), 344–354.

[24] Tuller, T., Waldman, Y. Y., Kupiec, M., & Ruppin, E. 2010b. Translation eficiency is determined by both codon bias and folding energy. *Proc. Natl. Acad .Sci. USA*, **107**(8), 3645–3650.

[25] von Heijne, G. 2006. Membrane-protein topology. *Nat. Rev. Mol. Cell Biol.*, **7**, 909–918.

[26] y Terán-Romero, L. Mier, Silber, M., & Hatzimanikatis, V. 2010. The origins of time-delay in template biopolymerization processes. *PLoS Comput. Biol.*, **6**(4), e1000726.

[27] Zavala, E., & Marquez-Lago, T. T. 2014. Delays Induce Novel Stochastic Effects in Negative Feedback Gene Circuits. *Biophys. J.*, **106**(2), 467–478.